

DEVICE AND METHOD FOR VOICE ACTIVITY DETECTION

Field of the invention

The present invention relates to a device, a mobile apparatus incorporating
5 the device, an accessory therefor and a method for voice activity detection,
particularly in a mobile telephone, using the directional sensitivity of a microphone
system and exploiting the knowledge about the voice source's orientation in space.
The device assists the existing voice activity detection to achieve higher sensitivity
and requiring less processor power.

10

State of the art

Voice activity detectors are used e.g. in mobile phones to enhance the
performance in certain situations. The most common way to construct a voice
activity detector is to look at the levels of the sub-bands of the incoming signal.
15 Then the background noise level and the speech level are estimated and compared
with a threshold to determine whether speech is present or not. An example of a
voice activity detector is disclosed in U.S. patent 6,427,134.

For instance in noisy environments it is hard to make a uniform parameter
set-up for the voice activity detector. Therefore several voice activity detectors are
20 needed, trimmed to the specific cases. For example in some modules you need to be
sure that if there is speech it should be detected (echo canceller), but in other cases
it is better to indicate no speech if the signal to noise ratio level is too low. The
plurality of voice activity detectors put a load on the digital signal processors that
have to take care of performing the various voice activity detection algorithms.

25

Summary of the invention

An object of the present invention is to complement existing voice activity
detection taking into account the direction of the source of the sound.

In a first aspect, the invention provides a device for voice activity detection
30 comprising a sound signal analyser arranged to determine whether a sound signal
comprises speech.

According to the invention, the device further comprises
a microphone system arranged to discriminate sounds emanating from sources
located in different directions from the microphone system, so that sounds only
35 emanating from a range of directions are included as signals possibly containing
speech.

Suitably, the range of directions is directed in the direction of an intended

elements separated a distance and located on a line directed in the direction of an intended user's mouth.

The range of directions may be defined as all sounds falling inside a cone with a cone angle α , wherein $10^\circ < \alpha < 30^\circ$, and preferably, α is approximately 25° .

5 In another embodiment, the microphone system comprises three microphone elements separated a distance and located in a plane directed in the direction of an intended user's mouth.

Suitably, two of said three microphone elements are separated a distance and located on a line directed perpendicular to the direction of an intended user's mouth.

10 In another embodiment, the microphone system comprises four microphone elements located such that the fourth microphone is not located in the same plane as the three others.

The microphone elements may be directional with a pattern having maximal sensitivity in the direction of an intended user's mouth.

15 In still a further embodiment, the microphone system comprises one directional microphone element together with one or more other microphone elements to remove the uncertainty in the direction of the sound source. The directional microphone element may be used to measure the sound pressure level relative to the other microphone element.

20 In a second aspect, the invention provides a mobile apparatus comprising a device as mentioned above.

Suitably, the microphone elements are located at the lower edge of the apparatus.

25 In one embodiment, a plurality of microphone elements are located at the lower edge of the apparatus and at least one further microphone element is located at a distance from the lower edge.

The mobile apparatus may be a mobile radio terminal, e.g. a mobile telephone, a pager, a communicator, an electric organiser or a smartphone.

30 In a third aspect, the invention provides an accessory for a mobile apparatus comprising a microphone system as mentioned above.

Suitably, the direction of the range of directions is adjustable.

The accessory may be a hands-free kit or a telephone conference microphone.

35 In a fourth aspect, the invention provides a method for voice activity detection, including the steps of:
 receiving sound signals from a microphone system arranged to discriminate sounds emanating from sources located in different directions from the microphone system;
 determining the direction of the sound source causing the sound signals;
 if the sounds emanate from a first range of directions, further analyse

the sound to determine whether the sound signal comprises speech;
but if the sounds emanate from a second, different range of directions decide that
the sound signal does not comprise speech.

Suitably, the first range of directions is directed in the direction of an
5 intended user's mouth.

The first range of directions may be defined as all sounds falling inside cone
with a cone angle α , wherein $10^\circ < \alpha < 30^\circ$, and preferably α is approximately 25° .

In one embodiment, the microphone system comprises at least two
microphone elements located at a distance from each other and located on a line
10 directed in the direction of an intended user's mouth, said two microphone elements
being separated a distance d , wherein the direction to the sound source θ is
calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

where

15 Δt is the time difference between the sounds from the two microphone elements,
 v is the velocity of sound.

In another embodiment, one directional microphone element is used together
with one or more other microphone elements to remove the uncertainty in the
direction of the sound source.

20 The directional microphone element may be used to measure the sound
pressure level relative to the other microphone element.

The invention is defined in the attached independent claims 1, 12, 16, and 20,
while preferred embodiments are set forth in the dependent claims.

25 Brief description of the drawings

The invention will be described below in greater detail with reference to the
accompanying drawings, in which:

fig. 1 is a perspective view of a mobile phone incorporating the present
invention, and

30 fig. 2 is a schematic drawing of the receiving angle of an embodiment of the
present invention.

Detailed description of preferred embodiments

As mentioned briefly in the introduction, many signal processing algorithms,
35 such as echo cancellation and background noise synthesis, used in phones and
hands-free kits are based on the fact that the user is speaking or not. For example
the speech codec is active when the near-end user is speaking and the background

voice activity detectors (VAD) to perform well. An error in the detection can result in artefacts or malfunctions caused by divergence of the algorithms or other problems.

Existing voice activity detectors are directed to determine whether speech is present or not in a sound signal. However, in fact not all speech is interesting or relevant, but only the user's speech. All other speech, e.g. in a noisy environment with several persons speaking, could be ignored and regarded as just noise.

The present inventor has realised that a microphone system having some kind of directional sensitivity could be used to discriminate sound emanating from different sources located in different directions. Sound not emanating from the user can be declared as non-speech, and those signals do not have to be analysed with the conventional voice activity detectors.

The existing voice activity detectors may be conventional and are only referred to as a sound signal analyser in this application.

Generally, a microphone system having some kind of directional sensitivity can be used. Fig. 1 shows an example with at least two separate microphone elements.

A general mobile telephone is indicated at 1. The invention is equally applicable to other devices such as mobile radio terminals, pagers, communicators, electric organisers or smartphones. The common feature is that voice activity detection is employed, e.g. in connection with communicating speech or receiving voice commands by means of speech recognition.

In the simplest version, the microphone system comprises two microphones 2a and 2b. Suitably, they are located on a line directed in the calculated direction of an intended user's mouth. Suitably, the microphone elements are located at the lower edge of the mobile apparatus 1.

Fig. 2 shows a schematic diagram of the calculation of the direction of the sound source, typically the user's mouth 3. In the case of two microphones, only the angle to the line on which the microphone elements are located can be determined. In other words, the direction of the sound source is on a cone with a cone angle θ . To calculate the angle θ , first a cross-correlation between the two signals from the microphones 2a and 2b is made. The maximum indicates the time difference Δt between the two microphones 2a and 2b. The distance between the two microphones 2a and 2b is e.g. 20 millimetres. The angle θ is calculated as

$$\theta = \arccos \frac{\Delta t \cdot v}{2 \cdot d}$$

Note that arccos is only defined for arguments between -1 and 1 . If the time

difference is negative, this means that the angle is greater than 90° and the sound emanates from behind the apparatus.

Suitably, the device is adapted to determine that all sounds with an angle θ less than a fixed angle α are emanating from the user. The threshold angle α may be
5 set within a range of e.g. 10° to 30° , suitably at 25° .

In the case of three microphones, the direction of the sound source can be further determined to be at two points (e.g. on the above cone). The three microphone elements are suitably located in a plane directed in the general direction of the user's mouth. In fig. 1 microphone elements 2b, 2c and 2d are a possible set-
10 up. The two microphone elements 2c and 2d at the front are located on a line perpendicular to the direction of the user's mouth, while the third microphone element 2b is located at the rear side.

In the case of four microphones (or more) detection of all direction angles may be calculated, provided that four microphone elements are located such that the
15 fourth microphone is not located in the same plane as the three others, e.g. on a tetrahedron. A possible set-up is two microphone elements 2c and 2d at the front on the lower edge, while a third microphone element 2b is located at the rear side, and a fourth microphone element 2e is located at the front at a distance from the lower edge.

20 A similar microphone arrangement may be used in an accessory to a mobile apparatus, such as a hands-free kit or a telephone conference microphone system intended to be placed on a table. Apart from the microphone elements the logic circuitry may be located in the main/mobile apparatus. In this case the reception angle of the microphone system can be adjustable. This is useful e.g. when the
25 microphone system is placed in a car, where the user can be seated either in the driver's seat or in the passenger's seat or even both the driver and the passenger may be speakers during the same call. The adjustment of the reception angle can be achieved mechanically or electronically, for example by beam forming or adaptation of the directional sensitivity of the microphone system.

30 To further enhance the sensitivity of the microphone system, directional microphone elements with a pattern having a maximum sensitivity in the direction of the user's mouth could be used.

In a further embodiment, one directional microphone element is used together with one or two other microphone elements (that may be non-directional).
35 The directional microphone element is used to measure the sound pressure level relative to the other(s), thus removing the uncertainty in the direction of the sound source. Various combinations of directional microphone elements and non-

performance. With the present invention only one voice activity detector may be necessary throughout the whole signal path. This will in turn reduce the computational complexity, decreasing the load on the digital signal processors as well as improving the performance. It is especially favourable in environments with
5 high background noise and noise with similar spectral properties as speech.

A person skilled in the art will realise that the invention may be realised with various combinations of hardware and software. The scope of the invention is only limited by the claims below.